APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Jan. 2004, p. 475–482 Vol. 70, No. 1
0099-2240/04/$08.00+0 DOI: 10.1128/AEM.70.1.475–482.2004

Species Differentiation of a Diverse Suite of *Bacillus* Spores by Mass Spectrometry-Based Protein Profiling

Danielle N. Dickinson,[1]* Myron T. La Duc,[2] William E. Haskins,[3] Igor Gornushkin,[1] James D. Winefordner,[1] David H. Powell,[1] and Kasthuri Venkateswaran[2]

Department of Chemistry[1] and McKnight Brain Institute,[3] University of Florida, Gainesville, Florida 32611, and NASA Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109[2]

Received 17 July 2003/Accepted 7 October 2003

In this study, we demonstrate the versatility of matrix-assisted laser desorption ionization–time-of-flight mass spectrometry (MALDI-TOFMS) protein profiling for the species differentiation of a diverse suite of *Bacillus* spores. MALDI-TOFMS protein profiles of 11 different strains of *Bacillus* spores, encompassing nine different species, were evaluated. *Bacillus* species selected for MALDI-TOFMS analysis represented the spore-forming bacterial diversity of typical class 100K clean room spacecraft assembly facilities. A one-step sample treatment and MALDI-TOFMS preparation were used to minimize the sample preparation time. A library of MALDI-TOFMS spectra was created from these nine *Bacillus* species, the most diverse protein profiling study of the genus reported to date. Linear correlation analysis was used to successfully differentiate the MALDI-TOFMS protein profiles from all strains evaluated in this study. The MALDI-TOFMS protein profiles were compared with 16S rDNA sequences for their bacterial systematics and molecular phylogenetic affiliations. The MALDI-TOFMS profiles were found to be complementary to the 16S rDNA analysis. Proteomic studies of *Bacillus subtilis* 168 were pursued to identify proteins represented by the biomarker peaks in the MALDI-TOFMS spectrum. Four small, acid-soluble proteins (A, B, C, and D), one DNA binding protein, hypothetical protein ymf J, and four proteins associated with the spore coat and spore coat formation (coat JB, coat F, coat T, and spoIVA) were identified. The ability to visualize higher-molecular-mass coat proteins (10 to 25 kDa) as well as smaller proteins (<10 kDa) with MALDI-TOFMS profiling is critical for the complete and effective species differentiation of the *Bacillus* genus.

Rapid, sensitive, and selective microbial detection and identification at the species and strain level are necessary to differentiate between viable pathogenic and nonpathogenic microbial species. The development of technology to accomplish this level of distinction for microbial species would have a significant impact on occupational and health care, homeland defense, and environmental monitoring. For over a century, microbial identification techniques have depended on conventional culture-based methods that characterize phenotypic differences and rely on biochemical and morphological tests. These methods are time-consuming and laborious, and the results are often subjective (38, 41). In order to overcome the problems involved with phenotypic characterization, 16S rRNA analysis has been used for decades to more accurately define the phylogenetic affiliation of the given test microorganism (29). However, being highly conserved, the 16S rRNA molecule at times cannot differentiate closely related microbial species (41, 43). Therefore, alternative biomarkers (44) or a suite of protein profiling methods would be useful in order to effectively differentiate closely related microbial species.

Matrix-assisted laser desorption ionization–time-of-flight mass spectrometry (MALDI-TOFMS)-based microbial detection technology has been evaluated to rapidly detect and discriminate microbial species. MALDI-TOFMS is well suited for this task due to its rapid analysis time (<1 min), low sample requirement (<2.5 μl), sensitivity, reproducibility, and resolving power. Analysis of whole bacterial cells and spores with this technique has given rise to unique protein fingerprints that can be used for identification at the species and strain level (5, 7, 9, 19, 20, 25, 26, 40). Vegetative cells generally produce a relatively large number of biomarker proteins that can be used for subsequent pattern recognition or correlation analysis (1, 16, 17). In contrast, the extraction of proteins from spores has been more challenging, giving only a limited number of biomarker peaks compared to results for their vegetative counterparts (8, 14, 35). Various sample pretreatments including infrared laser irradiation (34, 39), corona plasma discharge (4, 14), sonication (14), and the addition of 5% trifluoroacetic acid (34) or 1 M HCl (15) have been used to successfully extract proteins from spores. However, in most cases, these treatments required longer sample preparation times, and the visualization of peaks above 10 kDa is limited.

A majority of the MALDI-TOFMS research directed at spore detection has focused on only a few *Bacillus* species. These include *Bacillus anthracis* and its closely related species *B. thuringiensis* and *B. cereus* (43); *B. atrophaeus* (formally called *B. globigii*) (27), an anthrax surrogate; and *B. subtilis*, whose genome is completely sequenced and whose molecular biology has been thoroughly examined (8, 14, 15, 21, 32). Very little research has been done on other *Bacillus* species which naturally occur in the environment. The genus *Bacillus* is one of the largest, most ubiquitous, genera of bacteria, containing 65 valid species with new species being continually described (31). The nonpathogenic *Bacillus* spores are the most likely

* Corresponding author. Mailing address: University of Florida, Department of Chemistry, P.O. Box 117200, Gainesville, FL 32611-7200. Phone: (352) 392-2607. Fax: (352) 392-4651. E-mail: danielled@chem.ufl.edu.

TABLE 1. List of *Bacillus* species evaluated in this study

Organism	Strain	Source[a]	Remarks
B. atrophaeus	9372	ATCC	Surrogate to *B. anthracis*
B. licheniformis	14580	ATCC	Most predominant species in clean room facilities
B. licheniformis	KL-196	JPL-SAF	Class 100K clean room floor, JPL
B. megaterium	14581	ATCC	
B. mojavensis	51516	ATCC	
B. odysseyi	PTA-4399	KSC, SAEF-II	Mars Odyssey spacecraft surface
B. psychrodurans	VSE1-06	KSC, PHSF	Mars Exploration Rover assembly facility air particles
B. pumilus	7061	ATCC	Second-most predominant species in clean room facilities
B. subtilis	168	University of Arizona	Genome fully sequenced
B. subtilis	6051	ATCC	Type species of *Bacillus* genus
B. thuringiensis	10792	ATCC	Insecticide-producing bacteria and phylogenetically inseparable from *B. anthracis*

[a] SAF, spacecraft assembly facility; SAEF-II, spacecraft assembly and encapsulation facility-II; PHSF, payload hazardous service facility; JPL, Jet Propulsion Laboratory; KSC, Kennedy Space Center.

sources of interference for any detection technique and have the highest potential to produce false positives.

To demonstrate the versatility of MALDI-TOFMS protein profiling for the identification of a variety of spores, a subset of *Bacillus* species isolated from various NASA spacecraft assembly facilities (classes 10 to 100K clean rooms) was used in this study. A one-step sample treatment and MALDI-TOFMS preparation were used to rapidly obtain spectra with a wide range of protein biomarkers, including several higher-molecular-mass (10 to 25 kDa) protein species. A library of MALDI-TOFMS spectra was created from 11 different spores of *Bacillus* species, which encompassed nine different species, the most diverse study of the genus reported to date. Linear correlation analysis was used to identify all *Bacillus* species evaluated. The results obtained from MALDI-TOFMS protein profiling of these *Bacillus* spores were compared with 16S rDNA sequences for their bacterial systematics and molecular phylogenetic affiliations.

MATERIALS AND METHODS

Bacterial strains. *Bacillus* strains used in this study and their sources are listed in Table 1. Eleven strains consisting of nine *Bacillus* species were studied in which the type strains of *B. atrophaeus*, *B. licheniformis*, *B. megaterium*, *B. mojavensis*, *B. thuringiensis*, *B. pumilus*, and *B. subtilis* were procured from the American Type Culture Collection (ATCC; Manassas, Va.). *B. subtilis* 168 was received as a gift from Wayne Nicholson (University of Arizona). *B. odysseyi*, *B. licheniformis* KL-196, and *B. psychrodurans* were isolated from several NASA spacecraft and assembly facility surfaces. One additional strain each of *B. licheniformis* and *B. subtilis* was included in the study to compare the MALDI-TOFMS profiles of the same species. Bacterial isolation procedures from spacecraft and assembly facility surfaces were described elsewhere (22, 42). The identity of the test organisms was determined based on 16S ribosomal DNA (rDNA) sequencing for the environmental isolates, whereas for the ATCC strains, sequences available in the GenBank database were used (23). The 16S ribosomal DNA (rDNA) sequences of the environmental isolates have been deposited in the GenBank nucleotide sequence database.

Sporulation of *Bacillus* isolates. A nutrient broth sporulation medium was used to produce spores (28, 36). A single purified colony of the strain to be sporulated was inoculated into the nutrient broth sporulation liquid medium. After 1 to 3 days of incubation at 32°C under shaking conditions, cultures were examined by phase-contrast microscopy to determine the level of sporulation. Microcosms that attained >99% of spores were further purified to remove vegetative cells or cell debris as previously reported (28). The purified spores were suspended in sterile deionized water and stored at 4°C in glass tubes until analyzed. Before the analysis, spore suspensions were adjusted to give an optical density of 0.6 at 600 nm (OD_{600}), which resulted in suspensions that were between 10^8 to 10^9 spores/ml.

16S rDNA sequencing. Purified genomic DNA (18) from liquid-grown cultures was quantified and ~10 ng of DNA was used as the template for PCR amplification. Universal primers (Bact 11 and 1492) were used to amplify the 1.5-kb 16S rDNA fragment as per protocols established by Ruimy et al. (33). Amplicons generated were purified with Qiagen columns (Valencia, Calif.) and sequenced as described elsewhere (22, 42). The phylogenetic relationships of organisms covered in this study were determined by comparison of individual 16S rDNA sequences to other existing sequences in the public database (GenBank; http://www.ncbi.nlm.nih.gov/). Evolutionary trees were constructed with the PAUP program (37).

Sample preparation for mass spectrometry. A saturated matrix solution was prepared by dissolving 20 mg of ferulic acid into a 1-ml solution of 30% acetonitrile and 40% formic acid. This solvent system was selected due to the higher signal-to-noise ratio, consistent crystallization, and better ability to differentiate across the various bacterial species evaluated in this study. This effect is due to a combination of an increased number of biomarker peaks and the higher-molecular-weight range of these peaks in the spectra (D. N. Dickinson, W. E. Haskins, S. H. Powell, J. D. Winefordner, M. T. La Duc, and K. Venkateswaran, Proc. 52nd Annu. Meet. Am. Soc. Mass Spectrom., abstr. AO32385, 2003, and D. N. Dickinson, D. H. Powell, J. D. Winfornder, M. J. Kempf, and K. Venkateswaran, Proc. 51st Annu. Meet. Am. Soc. Mass Spectrom., abstr. AO20728, 2002). A 2.5-μl aliquot of the spore suspension (OD_{660}, 0.6) was added to 22.5 μl of the matrix solution. This mixture was vortexed briefly, and then 1 μl of the sample containing both spores and matrix compound was removed and spotted on a SCOUT26 MALDI plate (Bruker Daltonics, Billerica, Mass.). Spots were allowed to air dry. No further treatments were applied to the spots once they were dried. Two spots were prepared for each sample mixture. Sample preparation required only a few minutes per sample.

Mass spectrometry analysis. MALDI-TOFMS analysis was preformed on a Bruker Daltonics Reflex II mass spectrometer retrofitted with delayed extraction. The instrument was operated in the linear mode. A nitrogen laser (337 nm) pulsed at a frequency of 5 Hz irradiated the sample. Spectra were obtained in the positive ion mode with a delay time of 50 ns. The acceleration voltage was 20 kV. An ion deflector was used to deflect low-mass ions that would saturate the detector, and for all experiments the deflector was set at 3,000 Da. The laser intensity was adjusted to just above the threshold for ion formation for each sample. The instrument was calibrated daily by external calibration with a mixture of bovine insulin and equine cytochrome *c*. All spectra represent the accumulation of 50 laser shots. Ten spectra were collected from each spot on the MALDI plate. A total of 20 spectra were collected per spore sample.

Spectral processing and statistical methodology. Prior to statistical analysis, each spectrum was baseline corrected and smoothed according to a ten-point Savitzky-Golay smoothing algorithm. Normalized spectra were converted into ASCII files for statistical processing. Because linear correlation is invariant with respect to a linear transformation of spectra, the relative, not absolute, intensities were important for correlation analysis. Statistical analysis of the data was performed by using linear correlation software developed in-house with Visual Basic 6.0 (10–12). Spectra from the mass spectrometer were imported into the software as ASCII files, and libraries were created from the average of the 20 spectra collected per sample (10 spectra per spot). Correlation analysis was performed

TABLE 2. The 16S rDNA sequence similarities (%) for *Bacillus* species in this study

Organism and accession no.	*B. atrophaeus* X60607	*B. licheniformis* AF387515	*B. licheniformis* X68416	*B. megaterium* X60629	*B. mojavensis* AB021191	*B. odysseyi* AF526913	*B. psychrodurans* VSE1 06	*B. pumilus* AB020208	*B. subtilis* 168 rrnA	*B. subtilis* X60646	*B. thuringiensis* X55062
B. atrophaeus X60607	100										
B. licheniformis AF387515	96.9	100									
B. licheniformis X68416	98.5	98.3	100								
B. megaterium X60629	94.4	92.7	94.1	100							
B. mojavensis AB021191	99.3	96.7	98.4	94.1	100						
B. odysseyi AF526913	92.0	90.1	91.5	93.4	91.8	100					
B. psychrodurans VSE1 06	91.8	90.5	91.5	92.9	91.5	95.4	100				
B. pumilus AB020208	97.6	94.9	96.3	94.3	96.9	91.8	92.4	100			
B. subtilis 168 rrnA	99.4	96.9	98.6	94.1	99.7	91.6	91.4	97.2	100		
B. subtilis X60646	99.3	96.7	98.3	94.1	99.6	91.6	91.2	96.9	99.8	100	
B. thuringiensis X55062	95.2	92.9	94.2	94.7	94.3	92.8	92.0	94.3	94.2	94.3	100

on a point-to-point basis based on the following equation for the correlation coefficient r:

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2}\ \sqrt{\sum (y_i - \bar{y})^2}}$$

where $\bar{x}$ is the mean of x_i's and $\bar{y}$ is the mean of y_i's. x_i's and y_i's are the intensities at the i-th pixel of the detector ($i = 1 \ldots N$); the x_i's belong to an analyzed spectrum, and the y_i's belong to one of the library spectra. The spectrum consisting of x_i's is correlated against each spectrum in the library (different sets of y_i's), and the closest match with the highest correlation coefficient indicated a similarity of this spectrum with the corresponding library spectrum. On the other hand, the difference between this and other correlation coefficients signified spectral dissimilarities. To quantify the level of significance of these differences, a simple t test was applied. Student t values were calculated differently depending on whether the two distributions had the same or different variances. To check this, an F test (where F denotes the ratio of the variances) was applied as the basis of t values. The probabilities that two distributions of correlation coefficients had different means were calculated.

A reference library, comprising the average spectrum created from the 20 spectra collected for each spore sample, was produced for all of the 11 species evaluated in this study. The individual spectra and the average spectrum obtained from the 11 strains were then compared to the MALDI-TOFMS profiles of the various spores stored in the library to elucidate the bacterial speciation. To evaluate the reproducibility of the technique, a separate set of MALDI-TOFMS spectra was collected and averaged from all of the strains in this study. The averages of these separate analyses were compared with the library spectra. Additionally, to address batch-to-batch variability, *B. subtilis* 168 spore cultures prepared at different times over the course of 2 years were analyzed and compared to the library spectra.

Proteomic analysis. The MALDI protein extract from the spores of *B. subtilis* 168 were subjected to proteomic analysis. A 2.5-μl aliquot of the spore suspension (OD_{660}, 0.6) was diluted in 22.5 μl of 40% formic acid–30% acetonitrile. This sample was vortexed briefly and then centrifuged for 5 min at 9,600 × *g*. The supernatant (MALDI extract) was removed and placed in a clean microcentrifuge vial. The solvent was removed with a Speed Vac concentrator, and the sample was reconstituted in a solution of 50 mM ammonium bicarbonate. Trypsin was added at a 50:1 ratio to the sample, and the proteins were digested overnight at 37°C. Tryptic peptides were analyzed by capillary liquid chromatography-tandem mass spectrometry (CLC-MS^2) with a system similar to that described elsewhere (13). Sequence information was obtained for tryptic peptides via collision-induced dissociation. The mass-to-charge (m/z) ratios of the precursor ion and product ions for each tryptic peptide were searched against the National Center for Biotechnology Information protein database by using the Sequest (45) and Mascot (30) algorithms for protein identification.

RESULTS AND DISCUSSION

Incidence of spore-forming microbes from spacecraft and associated environment. Among several hundred aerobic spore-forming bacteria isolated from several spacecraft and associated facility surfaces, >90% of the isolates were found to be phylogenetically affiliated to the members of the genus *Bacillus* (22, 23, 42). *B. licheniformis* (25%) and *B. pumilus* (16%) were the most prevalent *Bacillus* species isolated. Since *B. licheniformis* was the most prevalent *Bacillus* species in the environment and *B. subtilis* is the type species of the *Bacillus* genus, multiple strains of these species were included in this study. *Bacillus* species selected in this study for MALDI-TOFMS analysis represent the spore-forming bacterial diversity of typical class 100K clean room facilities (22, 23, 42). In order to avoid confusion about the identity of the bacterial species, wherever possible, authentic type strains were procured from the culture collection and used. Phenotypically, all tested *Bacillus* species fall in group II except *B. psychrodurans* and *B. odysseyi*, which are in group IV (31). Group II includes aerobic *Bacillus* species that produce acid from a variety of sugars including glucose and whose spores are ellipsoidal and do not swell the mother cell. Group IV *Bacillus* species are also aerobic; however, they do not produce acids from sugars, and even though they also produce ellipsoidal spores, they swell the mother cell. The phenotypic group II *Bacillus* species are genotypically grouped as rRNA group I and the phenotypic group IV *Bacillus* species studied here are considered rRNA group II (3). As the *Bacillus* species of other rRNA groups were not isolated from class 100K clean room facilities (22–24, 42), we restricted the characterization of the species by MALDI-TOFMS to 11 members of these two rRNA groups.

Molecular phylogeny of spore-forming microbes. The sequence similarities based on 16S rDNA sequences of the various *Bacillus* species tested are shown in Table 2. These sequences were either obtained from the GenBank database or were sequenced in previous studies (22–24, 42). The similarities in 16S rDNA nucleotide sequences between the tested

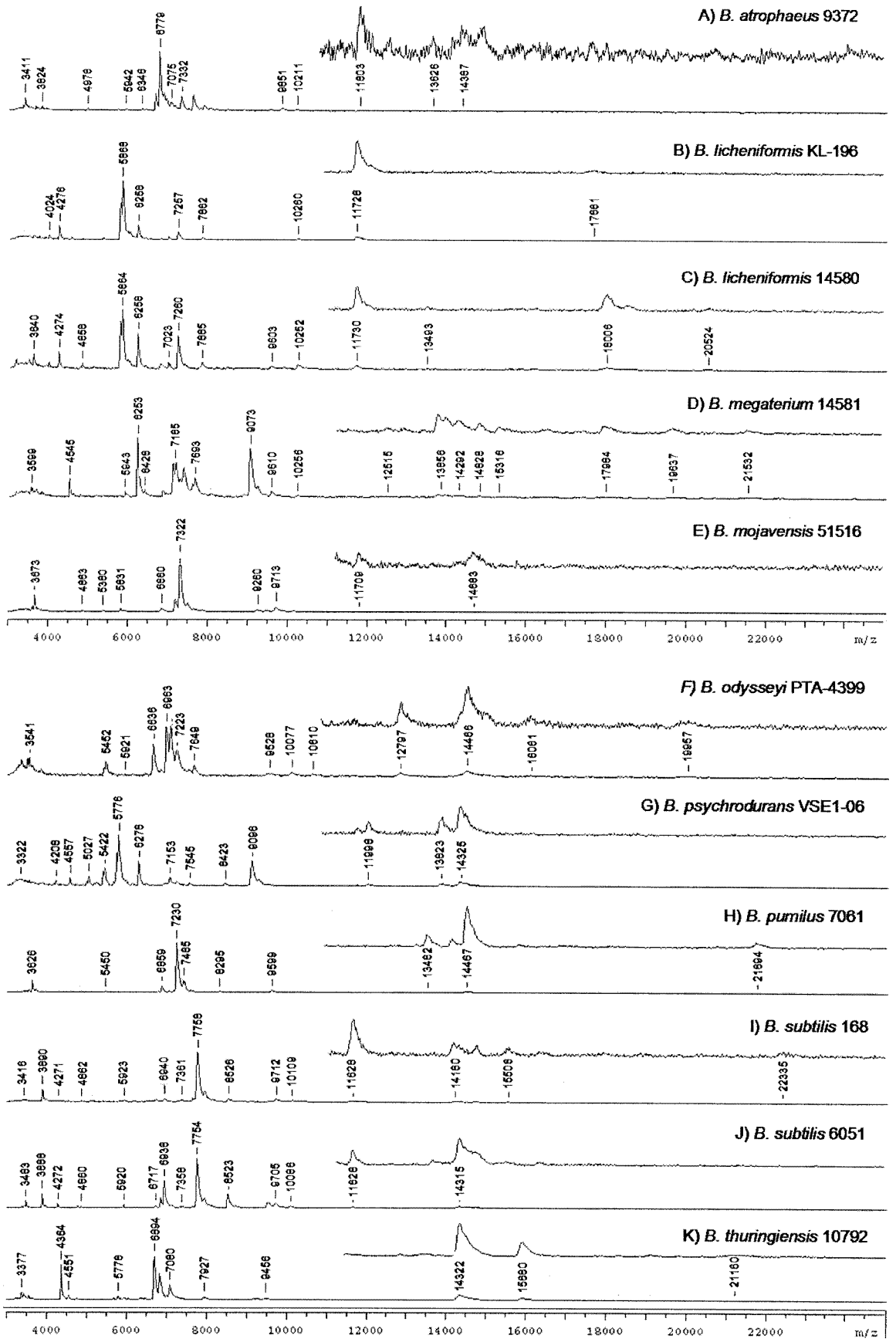

FIG. 1. MALDI-TOFMS protein profiles of the 11 *Bacillus* species analyzed in this study. The mass range depicted is from *m*/*z* 3,000 to 25,000. The higher-molecular-mass region from *m*/*z* 9,500 to 25,000 is amplified four times (see inset of each spectrum) in order to visualize the higher-molecular-weight peaks that are present but are at much lower abundance in the samples.

Bacillus species, recognized by GenBank BLAST searches, were between 91 and 99%. A sequence variation of ~9% was found between rRNA groups I and II *Bacillus* species. A very high sequence variation within a well-described genus is not uncommon. Further analyses indicated that *B. atrophaeus* shares a close phylogenetic relationship with several *Bacillus* species such as *B. mojavensis*, *B. pumilus*, and *B. subtilis* (>97.5%). Similarly, *B. licheniformis* wild-type strain KL-196 and *B. mojavensis*, as well as two *B. subtilis* strains tested in this study, showed >98% 16S rDNA sequence similarities. Such high 16S rDNA sequence similarities were also noticed (>99%) in the case of the two *B. subtilis* strains sequenced and *B. mojavensis.* This finding clearly shows that 16S rDNA sequence analysis was not useful in differentiating these closely related species of the genus *Bacillus.* The species identities of all these strains were confirmed by DNA-DNA hybridization (data not shown). The two strains of *B. licheniformis* and *B. subtilis* showed >70% DNA-DNA hybridization dissociation values and exhibited >98.5% 16S rDNA sequence similarities. When all these species were grouped together, the maximum-likelihood-based phylogenetic tree showed two major clusters (data not shown). One cluster consists of *B. megaterium, B. odysseyi*, *B. psychrodurans*, and *B. thuringiensis*, where the spores of these species contained an additional structure called exosporium around the spore outer coat. The second cluster formed by the other species tested does not contain an exosporium.

MALDI-TOFMS profiles. A representative spectrum from each *Bacillus* species analyzed in this study is shown in Fig. 1. The mass spectrum depicted is from *m*/*z* 3,000 to 25,000. The higher *m*/*z* region from 9,500 to 25,000 is amplified (see inset of each spectrum) in order to visualize the less abundant peaks present at higher *m*/*z* regions. The observation of proteins at a higher *m*/*z* region is seldom reported in other MALDI-TOFMS analyses of whole spores (8, 14, 15, 34, 39) We hypothesize that the appearance of large proteins with high *m*/*z* values is due to optimization of the solvent system (Dickinson et. al., Proc. 52nd Annu. Meet. Am. Soc. Mass Spectrom., and Dickinson et al., Proc. 51st Annu. Meet. Am. Soc. Mass Spectrom.) used in this study. The solvent system was optimized with respect to signal-to-noise ratios, reproducibility, and the number of proteins extracted. It was found that preparations containing formic acid were more effective at extracting proteins from spores compared to standard MALDI preparations containing trifluoroacetic acid. As the formic acid concentration increases, so do the number and molecular weight range of proteins extracted from the spores. The solvent system chosen in this study represents the optimum balance between maximizing the protein extraction and maintaining the homogeneity and reproducibility of the MALDI crystal formation. Significantly, this method provides more confident identification of the various strains of bacteria at the species level.

From the spectra, we were unable to identify an obvious *Bacillus*-ubiquitous biomarker with the sample preparation protocol adapted in this study. A peak at *m*/*z* 14,500 was present in all of the spore spectra obtained except for that of the *B. licheniformis* ATCC 14580 type strain and its wild-type strain KL-196. The absence of a genus-specific biomarker may be due to the extraction protocol used in this study, posttranslational modifications of proteins that may differ between the strains, or the need for more sophisticated spectral comparisons of the different species. All of the spores have a group of peaks in the *m*/*z* region between 6,500 and 8,000. *B. licheniformis* ATCC 14580, *B. licheniformis* KL-196, *B. psychrodurans*, *B. odysseyi*, and *B. megaterium* all have an additional group of peaks between *m*/*z* 5,000 to 6,500 that is not observable in the other spectra. It was challenging to obtain good spectra from the *B. odysseyi* samples as shown by the lower signal-to-noise ratio in the spectra. This could be a result of glycoproteins present in the exosporium layers. Glycoproteins can be difficult to analyze due to the poor ionization of the sugar moieties in the MALDI process and the inherent heterogeneity of glycosolations. An expected result was the level of similarity observed between the strains of the same species. *B. licheniformis* ATCC 14580 type strain (Fig. 1C) and its wild-type strain KL-196 (Fig. 1B) and *B. subtilis* 168 (Fig. 1I) and ATCC 6051 (Fig. 1J) have very comparable MALDI-TOFMS profiles upon visual inspection. The spectra for the *B. licheniformis* pair are very similar except for a difference in the intensity of the *m*/*z* 7,260 peak and the presence of different higher-molecular-mass species in *B. licheniformis* 14580. The *B. subtilis* strains exhibit the same pattern in that there is a difference in peak intensity for the peak at *m*/*z* 6,936 and variation in the masses observed above *m*/*z* 10,000. This observation supports the theory that it is important to examine a wide variety of *Bacillus* spores before assigning definitive genus-, species-, and strain-specific protein biomarkers.

Linear correlation analysis provided a means of statistical comparison of the spectra. Correlation values close to 1 indicate that the fingerprint patterns of two organisms are very similar. Table 3 shows the linear correlation values for the MALDI-TOFMS spectra of the various *Bacillus* species evaluated compared to the library spectra. Each of the 20 individual spectra from each species was searched against the user-generated average library spectra. All individual spectra were successfully identified as their corresponding species and strain. These results were verified by applying Student's *t* test to the data. Using the *t* test, we confirmed that we were able to differentiate all the species studied at the 95% confidence level. Figure 2 shows the correlation results of the 20 individual *B. atrophaeus* spectra when they were searched against the library spectra. The *y* axis represents the linear correlation values obtained and the *x* axis represents the first to fifth ranks (hits) from the library. At each rank, the standard deviation of the measurement across the 20 spectra is represented by the error bars. The graph demonstrates that for rank 1 (*B. atrophaeus*), we have very high correlation values (0.96 ± 0.02) and that for the next best hit, *B. thuringiensis*, the correlation values are much lower (0.51 ± 0.02). Since none of the correlation values approaches the *B. atrophaeus* hit, we can confirm the differentiation of *B. atrophaeus* from all of the other strains in the library. The linear correlation method applied here also allows for the differentiation of the species whose MALDI-TOFMS profiles are almost indistinguishable upon visual observation, including the type strain and wild-type strains of *B. subtilis* and *B. licheniformis*. Figure 3 shows the correlation results of *B. subtilis* 168 versus the library spectra as described above. The second rank (or hit) is much closer than in the case of *B. atrophaeus*; the values are 0.98 ± 0.02 for the first rank and 0.86 ± 0.02 for the second rank. The second rank repre-

TABLE 3. Correlation values for MALDI-TOFMS spectra of various *Bacillus* species

Organism and strain	*B. atrophaeus* ATCC 9372	*B. licheniformis* ATCC 14580	*B. licheniformis* KL-196	*B. megaterium* ATCC 14581	*B. mojavensis* ATCC 51516	*B. odysseyi* PTA-4399	*B. psychrodurans* VSE1-06	*B. pumilus* ATCC 7061	*B. subtilis* 168	*B. subtilis* ATCC 6051	*B. thuringiensis* ATCC 10792
B. atrophaeus ATCC 9372	1										
B. licheniformis ATCC 14580	0.09	1									
B. licheniformis KL-196	0.03	0.90	1								
B. megaterium ATCC 14581	0.14	0.30	0.09	1							
B. mojavensis ATCC 51516	0.20	0.23	0.07	0.37	1						
B. odysseyi PTA-4399	0.35	0.15	0.05	0.26	0.21	1					
B. psychrodurans VSE1-06	0.04	0.45	0.42	0.35	0.05	0.16	1				
B. pumilus ATCC 7061	0.14	0.27	0.04	0.45	0.44	0.34	0.07	1			
B. subtilis 168	0.09	0.04	0.01	0.11	0.07	0.05	−0.01	0.02	1		
B. subtilis ATCC 6051	0.23	0.05	0.01	0.11	0.07	0.29	0.01	0.05	0.88	1	
B. thuringiensis ATCC 10792	0.52	0.06	0.03	0.07	0.05	0.33	0.08	0.09	0.02	0.13	1

sents *B. subtilis* ATCC 6051, the other *B. subtilis* strain in this study. With statistical treatment of the data, we are still able to differentiate the two strains at the 97% confidence interval. The very close correlation values of 0.88 ± 0.02 for the *B. licheniformis* pair and 0.86 ± 0.02 for the *B. subtilis* pair illustrate that close correlation values do indicate a relationship between the organisms. However, with statistical treatment of the data, we are able to obtain differentiation at the strain level in these two examples.

To ascertain the robustness of the technique, separate spectra collected and averaged from the same spore culture were examined. All 11 species were correctly identified by comparison to the library spectra (r = 0.85 to 0.98). This result was consistent whether the individual spectra or averages of individual spectra were used to search the library. In addition to the new preparations from the same culture, four batches of spores of *B. subtilis* 168, prepared at different times over the course of 2 years, were also compared against the library spectra. All of the *B. subtilis* 168 spores were correctly identified as the *B. subtilis* 168 from the library, regardless of the batch or storage time (r = 0.92 to 0.98).

Aligning the correlation results from the MALDI-TOFMS profiles (Table 3) with the 16S rDNA sequence analysis (Table 2) shows that the MALDI-TOFMS profiles are complementary to 16S rDNA analysis. Using MALDI-TOFMS profiles of these organisms, we are able to confidently differentiate all of the species studied, whereas there are several species including *B. subtilis* 168, *B. licheniformis*, *B. mojavensis*, and *B. atrophaeus* that 16S rDNA analysis is unable to differentiate at the

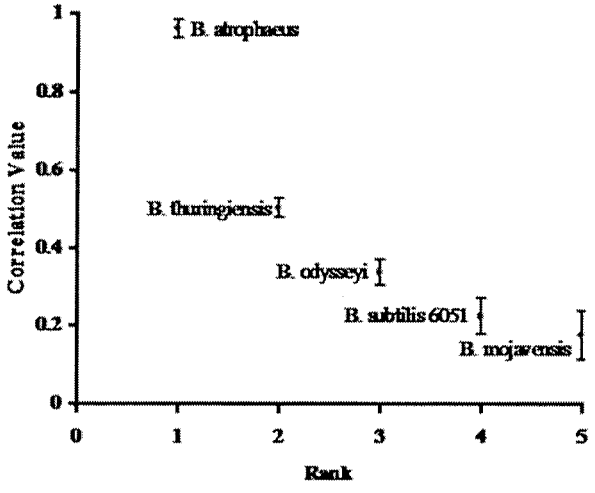

FIG. 2. Correlation results of the 20 individual *B. atrophaeus* ATCC 9372 spectra searched against the library. The *y* axis represents the linear correlation values obtained and the *x* axis represents the first to fifth ranks (hits) from the library. At each rank, the standard deviation of the measurement across the 20 spectra is represented by the error bars.

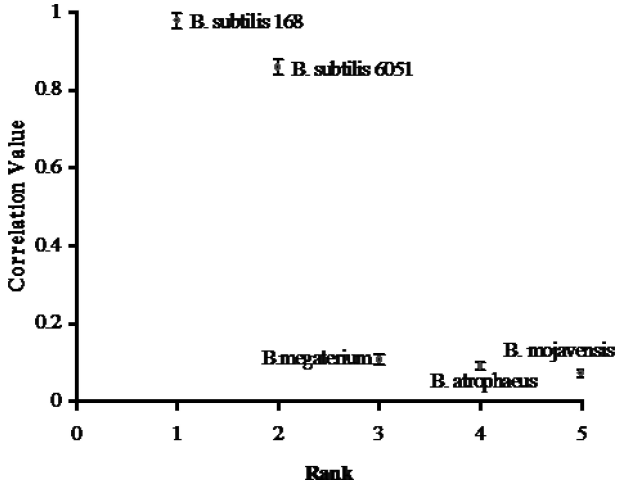

FIG. 3. Correlation results of the 20 individual *B. subtilis* 168 spectra searched against the library. The *y* axis represents the linear correlation values obtained and the *x* axis represents the first to fifth ranks (hits) from the library. At each rank, the standard deviation of the measurement across the 20 spectra is represented by the error bars.

TABLE 4. Proteins from *B. subtilis* 168 MALDI extract identified with proteomic analysis

Protein	Description[a]	Mascot or Xcorr score	% Coverage	Accession no.	Mol mass (Da)
CotT	Spore coat protein precursor	450	59.7	1075916	10,125
SspB	Spore protein; major beta-type SASP	203	61.2	16078040	6,975
HU	DNA binding protein	178	41.3	16079336	9,878
CotF	Spore coat protein	176	37.5	116958	18,714
SspA	Spore protein; major alpha-type SASP	150	62.3	16080009	7,066
YmfJ	Conserved hypothetical protein	114	56.5	16078751	9,642
CotJB	Spore coat peptide assembly protein	76	26.0	16077757	11,745
SspD	Spore protein; minor alpha/beta-type SASP	67	43.7	16078411	6,800
SspC	Spore protein; SASP	65	27.8	134238	7,753
SpoIVA	Coat morphogenesis sporulation protein	50	4.3	16079337	55,140

[a] SASP, small, acid-soluble protein.

species level. MALDI-TOFMS analysis of these species would allow for differentiation at the species level. Comparing the MALDI protein profiles with the phenotypic groupings was challenging due to the large diversity in the number and range of the peaks across the spectra for all of the species studied. In general, spores with an exosporium resulted in spectra that had more peaks over a broader range than the organisms without an exosporium. On average, the phenotypic group IV organisms had more peaks than the group II organisms, with the exception of *B. megaterium.* Other statistical methods, such as cluster analysis, could be applied to the data to allow us to visualize on the basis of protein profiles the relationships between the different species.

Proteomic analysis for biomarker identification. Since the genome of *B. subtilis* 168 is completely sequenced, this organism was selected for further proteomic studies towards identifying which proteins are represented by the biomarker peaks observed by MALDI-TOFMS. We were able to identify by CLC-MS^2 10 proteins in the extract of the spore sample of *B. subtilis* 168. The protein description, sequence coverage, molecular masses, and database searching scores (i.e., Sequest Xcorr or Mascot score) are shown in Table 4. Four small, acid-soluble proteins (A, B, C, and D), one DNA binding protein, hypothetical protein YmfJ, and four proteins associated with the spore coat and spore coat formation (CotJB, CotF, CotT, and SpoIVA) were identified.

Relating these proteins to the peaks observed in the MALDI-TOFMS profile for *B. subtilis* 168 was challenging. Proteins in the lower-molecular-mass region (below *m*/*z* 10,000) represent the small, acid-soluble proteins, the DNA binding protein, and smaller spore coat polypeptides processed from larger precursors. The higher *m*/*z* peaks may represent other processed and intact spore coat proteins such as CotJB at *m*/*z* 11,638 Da. The large peak at *m*/*z* 7,758 is the processed form of CotT, which starts as a 10-kDa protein in which the first 19 residues (termed the propeptide) are removed to leave behind an ~7,800-Da spore coat protein (2). The molecular masses listed in the protein database of four of the proteins identified (SpoIVA, YmfJ, CotF, and SspB) did not directly match with the *m*/*z* values of singly charged ions observed in the MALDI spectra. Separation of the proteins prior to proteomic analysis is required in order to reduce the complexity of the biomarker protein extract and confidently assign protein identifications made by CLC-MS^2 to peaks in the MALDI spectra. This is the first time to our knowledge that proteins associated with the spore coat have been identified from direct spore analysis by using a MALDI extract, as previous studies have only identified small, acid-soluble proteins found in the spore cortex as the source of the biomarker peaks. While small, acid-soluble proteins do allow for some species differentiation, they do not account for all the peaks observed, nor do they allow for differentiation at the species level as in the case of *B. thuringiensis* and *B. cereus* (5, 6, 15). Therefore, the ability to visualize and identify coat proteins as well as small, acid-soluble proteins in the spore is critical for complete and effective species differentiation of the *Bacillus* genus.

Protein profiling based on MALDI-TOFMS is a useful, rapid, and sensitive technology to differentiate spores from closely related microbial species. Although a standardized sample preparation protocol is required, it is obvious from the result that this technology is promising for species differentiation of a wide variety of bacterial spores. Additional optimization of the solvent system used for the MALDI-TOFMS analysis may extend the molecular mass range further and provide more biomarkers for subsequent proteomic analysis. Complete characterization of the protein biomarkers observed in this study is necessary to bring this technique into full fruition as a viable microbial analysis tool.

ACKNOWLEDGMENTS

Part of the research described in this publication was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

D. Dickinson thanks the NASA Graduate Student Research Program for support. We thank A. Driks, W. Nicholson, and M. Rasche for technical discussion and K. Buxbaum and D. Jan for encouragement.

REFERENCES

1. **Arnold, R. J., and J. P. Reilly.** 1998. Fingerprint matching of E-coli strains with matrix-assisted laser desorption ionization time-of-flight mass spectrometry of whole cells using a modified correlation approach. Rapid Commun. Mass Spectrom. **12:**630–636.
2. **Aronson, A. I., H. Y. Song, and N. Bourne.** 1989. Gene structure and precursor processing of a novel Bacillus subtilis spore coat protein. Mol. Microbiol. **3:**437–444.
3. **Ash, C., J. A. Farrow, M. Dorsch, E. Stackebrandt, and M. D. Collins.** 1991. Comparative analysis of *Bacillus anthracis*, *Bacillus cereus*, and related species on the basis of reverse transcriptase sequencing of 16S rRNA. Int. J. Syst. Bacteriol. **41:**343–346.
4. **Birmingham, J., P. Demirev, Y. P. Ho, J. Thomas, W. Bryden, and C. Fenselau.** 1999. Corona plasma discharge for rapid analysis of microorganisms by mass spectrometry. Rapid Commun. Mass Spectrom. **13:**604–606.

5. **Demirev, P. A., Y. P. Ho, V. Ryzhov, and C. Fenselau.** 1999. Microorganism identification by mass spectrometry and protein database searches. Anal. Chem. **71:**2732–2738.
6. **Demirev, P. A., J. Ramirez, and C. Fenselau.** 2001. Tandem mass spectrometry of intact proteins for characterization of biomarkers from Bacillus cereus T spores. Anal. Chem. **73:**5725–5731.
7. **Easterling, M. L., C. M. Colangelo, R. A. Scott, and I. J. Amster.** 1998. Monitoring protein expression in whole bacterial cells with MALDI time-of-flight mass spectrometry. Anal. Chem. **70:**2704–2709.
8. **Elhanany, E., R. Barak, M. Fisher, D. Kobiler, and Z. Altboum.** 2001. Detection of specific Bacillus anthracis spore biomarkers by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Rapid Commun. Mass Spectrom. **15:**2110–2116.
9. **Fenselau, C., and P. A. Demirev.** 2001. Characterization of intact microorganisms by MALDI mass spectrometry. Mass Spectrom. Rev. **20:**157–171.
10. **Gornushkin, I. B., A. Ruiz-Medina, J. M. Anzano, B. W. Smith, and J. D. Winefordner.** 2000. Identification of particulate materials by correlation analysis using a microscopic laser induced breakdown spectrometer. J. Anal. At. Spectrom. **15:**581–586.
11. **Gornushkin, I. B., B. W. Smith, H. Nasajpour, and J. D. Winefordner.** 1999. Identification of solid materials by correlation analysis using a microscopic laser-induced plasma spectrometer. Anal. Chem. **71:**5157–5164.
12. **Gornushkin, S. I., I. B. Gornushkin, J. M. Anzano, B. W. Smith, and J. D. Winefordner.** 2002. Effective normalization technique for correction of matrix effects in laser-induced breakdown spectroscopy detection of magnesium in powdered samples. Appl. Spectrosc. **56:**433–436.
13. **Haskins, W. E., Z. Q. Wang, C. J. Watson, R. R. Rostand, S. R. Witowski, D. H. Powell, and R. T. Kennedy.** 2001. Capillary LC-MS2 at the attomole level for monitoring and discovering endogenous peptides in microdialysis samples collected in vivo. Anal. Chem. **73:**5005–5014.
14. **Hathout, Y., P. A. Demirev, Y.-P. Ho, J. L. Bundy, V. Ryzhov, L. Sapp, J. Stutler, J. Jackman, and C. Fenselau.** 1999. Identification of *Bacillus* spores by matrix-assisted laser desorption ionization–mass spectrometry. Appl. Environ. Microbiol. **65:**4313–4319.
15. **Hathout, Y., B. Setlow, R. M. Cabrera-Martinez, C. Fenselau, and P. Setlow.** 2003. Small, acid-soluble proteins as biomarkers in mass spectrometry analysis of *Bacillus* spores. Appl. Environ. Microbiol. **69:**1100–1107.
16. **Holland, R. D., J. G. Wilkes, F. Rafii, J. B. Sutherland, C. C. Persons, K. J. Voorhees, and J. O. Lay.** 1996. Rapid identification of intact whole bacteria based on spectral patterns using matrix-assisted laser desorption/ionization with time-of-flight mass spectrometry. Rapid Commun. Mass Spectrom. **10:**1227–1232.
17. **Jarman, K. H., S. T. Cebula, A. J. Saenz, C. E. Petersen, N. B. Valentine, M. T. Kingsley, and K. L. Wahl.** 2000. An algorithm for automated bacterial identification using matrix-assisted laser desorption/ionization mass spectrometry. Anal. Chem. **72:**1217–1223.
18. **Johnson, J. L.** 1981. Genetic characterization, p. 34–51. *In* P. Gerhardt, R. G. E. Murray, R. N. Costilaw, E. W. Nester, W. A. Wood, N. R. Kreig, and G. B. Phillips (ed.), Manual of methods for general bacteriology. American Society for Microbiology, Washington, D.C.
19. **Krishnamurthy, T., and P. L. Ross.** 1996. Rapid identification of bacteria by direct matrix-assisted laser desorption/ionization mass spectrometric analysis of whole cells. Rapid Commun. Mass Spectrom. **10:**1992–1996.
20. **Krishnamurthy, T., P. L. Ross, and U. Rajamani.** 1996. Detection of pathogenic and nonpathogenic bacteria by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Rapid Commun. Mass Spectrom. **10:**883–888.
21. **Kunst, F., N. Ogasawara, I. Moszer, A. M. Albertini, G. Alloni, V. Azevedo, M. G. Bertero, P. Bessieres, A. Bolotin, S. Borchert, R. Borriss, L. Boursier, A. Brans, M. Braun, S. C. Brignell, S. Bron, S. Brouillet, C. V. Bruschi, B. Caldwell, V. Capuano, N. M. Carter, S. K. Choi, J. J. Codani, I. F. Connerton, N. J. Cummings, R. A. Daniel, F. Denizot, K. M. Devine, A. Dusterhoft, S. D. Ehrlich, P. T. Emmerson, K. D. Entian, J. Errington, C. Fabret, E. Ferrari, D. Foulger, C. Fritz, M. Fujita, Y. Fujita, S. Fuma, A. Galizzi, N. Galleron, S. Y. Ghim, P. Glaser, A. Goffeau, E. J. Golightly, G. Grandi, G. Guiseppi, B. J. Guy, K. Haga, J. Haiech, C. R. Harwood, A. Henaut, H. Hilbert, S. Holsappel, S. Hosono, M. F. Hullo, M. Itaya, L. Jones, B. Joris, D. Karamata, Y. Kasahara, M. KlaerrBlanchard, C. Klein, Y. Kobayashi, P. Koetter, G. Koningstein, S. Krogh, M. Kumano, K. Kurita, A. Lapidus, S. Lardinois, J. Lauber, V. Lazarevic, S. M. Lee, A. Levine, H. Liu, S. Masuda, C. Mauel, C. Medigue, N. Medina, R. P. Mellado, M. Mizuno, D. Moestl, S. Nakai, M. Noback, D. Noone, M. O'Reilly, K. Ogawa, A. Ogiwara, B. Oudega, S. H. Park, V. Parro, T. M. Pohl, D. Portetelle, S. Porwollik, A. M. Prescott, E. Presecan, P. Pujic, B. Purnelle, G. Rapoport, M. Rey, S. Reynolds, M. Rieger, C. Rivolta, E. Rocha, B. Roche, M. Rose, Y. Sadaie, T. Sato, E. Scanlan, S. Schleich, R. Schroeter, F. Scoffone, J. Sekiguchi, A. Sekowska, S. J. Seror, P. Serror, B. S. Shin, B. Soldo, A. Sorokin, E. Tacconi, T. Takagi, H. Takahashi, K. Takemaru, M. Takeuchi, A. Tamakoshi, T. Tanaka, P. Terpstra, A. Tognoni, V. Tosato, S. Uchiyama, M. Vandenbol, F. Vannier, A. Vassarotti, A. Viari, R. Wambutt, E. Wedler, H. Wedler, T. Weitzenegger, P. Winters, A. Wipat, H. Yamamoto, K. Yamane, K. Yasumoto, K. Yata, K. Yoshida, H. F. Yoshikawa, E. Zumstein, H. Yoshikawa, and A. Danchin.** 1997. The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. Nature **390:**249–256.
22. **La Duc, M. T., R. Kern, and K. Venkateswaran.** Microbial monitoring of spacecraft associated environments. Microb. Ecol., in press.
23. **La Duc, M. T., W. L. Nicholson, R. Kern, and K. Venkateswaran.** 2003. Microbial characterization of the Mars Odyssey spacecraft and its encapsulation facility. Environ. Microbiol. **10:**977–985.
24. **La Duc, M. T., M. Satomi, and K. Venkateswaran.** *Bacillus odysseyi* sp. nov. isolated from the Mars Odyssey spacecraft. Int. J. Syst. Evol. Microbiol., in press.
25. **Lay, J. O.** 2000. MALDI-TOF mass spectrometry and bacterial taxonomy. TRAC Trends Anal. Chem. **19:**507–516.
26. **Lay, J. O.** 2001. MALDI-TOF mass spectrometry of bacteria. Mass Spectrom. Rev. **20:**172–194.
27. **Nakamura, L. K.** 1989. Taxonomic relationship of black-pigmented *Bacillus subtilis* strains and a proposal for *Bacillus atrophaeus* sp. nov. Int. J. Syst. Bacteriol. **39:**295–300.
28. **Nicholson, W. L., and P. Setlow.** 1990. Sporulation, germination, and outgrowth, p. 391–450. *In* C. R. Harwood and S. M. Cutting (ed.), Molecular biological methods for Bacillus. John Wiley & Sons, Chichester, West Sussex, England.
29. **Pace, N. R., and T. L. Marsh.** 1985. RNA catalysis and the origin of life. Orig. Life Evol. Biosph. **16:**97–116.
30. **Perkins, D. N., D. J. Pappin, D. M. Creasy, and J. S. Cottrell.** 1999. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis **20:**3551–3567.
31. **Priest, F. G.** 1993. Systematics and ecology of *Bacillus*, p. 3–17. *In* A. L. Sonenshein, R. Losick, and J. A. Hock (ed.), *Bacillus subtilis* and other gram-positive bacteria: biochemistry, physiology, and molecular genetics. American Society for Microbiology, Washington, D.C.
32. **Ramirez, J., and C. Fenselau.** 2001. Factors contributing to peak broadening and mass accuracy in the characterization of intact spores using matrix-assisted laser desorption/ionization coupled with time-of-flight mass spectrometry. J. Mass Spectrom. **36:**929–936.
33. **Ruimy, R., V. Breittmayer, P. Elbaze, B. Lafay, O. Boussemart, M. Gauthier, and R. Christen.** 1994. Phylogenetic analysis and assessment of the genera *Vibrio*, *Photobacterium*, *Aeromonas*, and *Plesiomonas* deduced from small-subunit rRNA sequences. Int. J. Syst. Bacteriol. **44:**416–426.
34. **Ryzhov, V., J. L. Bundy, C. Fenselau, N. Taranenko, V. Doroshenko, and C. R. Prasad.** 2000. Matrix-assisted laser desorption/ionization time-of-flight analysis of Bacillus spores using a 2.94 micron infrared laser. Rapid Commun. Mass Spectrom. **14:**1701–1706.
35. **Ryzhov, V., Y. Hathout, and C. Fenselau.** 2000. Rapid characterization of spores of *Bacillus cereus* group bacteria by matrix-assisted laser desorption-ionization time-of-flight mass spectrometry. Appl. Environ. Microbiol. **66:**3828–3834.
36. **Schaeffer, P., J. Millet, and J. P. Aubert.** 1965. Catabolic repression of bacterial sporulation. Proc. Natl. Acad. Sci. USA **54:**704–708.
37. **Swofford, D.** 1990. PAUP v 3.0: phylogenetic analysis using parsimony. Illinois Natural History Survey, University of Illinois, Champaign, Ill.
38. **Towner, K. J., and A. Cockayne.** 1993. Molecular methods for microbial identification and typing. Chapman & Hall, London, England.
39. **Ullom, J. N., M. Frank, E. E. Gard, J. M. Horn, S. E. Labov, K. Langry, F. Magnotta, K. A. Stanion, C. A. Hack, and W. H. Benner.** 2001. Discrimination between bacterial spore types using time-of-flight mass spectrometry and matrix free infrared laser desorption and ionization. Anal. Chem. **73:**2331–2337.
40. **van Baar, B. L. M.** 2000. Characterisation of bacteria by matrix-assisted laser desorption/ionisation and electrospray mass spectrometry. FEMS Microbiol. Rev. **24:**193–219.
41. **Venkateswaran, K., D. P. Moser, M. E. Dollhopf, D. P. Lies, D. A. Saffarini, B. J. MacGregor, D. B. Ringelberg, D. C. White, M. Nishijima, H. Sano, J. Burghardt, E. Stackebrandt, and K. H. Nealson.** 1999. Polyphasic taxonomy of the genus *Shewanella* and description of *Shewanella oneidensis* sp. nov. Int. J. Syst. Bacteriol. **49:**705–724.
42. **Venkateswaran, K., M. Satomi, S. Chung, R. Kern, R. Koukol, C. Basic, and D. White.** 2001. Molecular microbial diversity of a spacecraft assembly facility. Syst. Appl. Microbiol. **24:**311–320.
43. **Yamada, S., E. Ohashi, N. Agata, and K. Venkateswaran.** 1999. Cloning and nucleotide sequence analysis of *gyrB* of *Bacillus cereus*, *B. thuringiensis*, *B. mycoides*, and *B. anthracis* and their application to the detection of *B. cereus* in rice. Appl. Environ. Microbiol. **65:**1483–1490.
44. **Yamamoto, S., and S. Harayama.** 1996. Phylogenetic analysis of *Acinetobacter* strains based on the nucleotide sequences of *gyrB* genes and on the amino acid sequences of their products. Int. J. Syst. Bacteriol. **46:**506–511.
45. **Yates, J. R., III, S. F. Morgan, C. L. Gatlin, P. R. Griffin, and J. K. Eng.** 1998. Method to compare collision-induced dissociation spectra of peptides: potential for library searching and subtractive analysis. Anal. Chem. **70:**3557–3565.